United Nations ECE/CES/2015/29

Economic and Social Council

Distr.: General
2 April 2015

Original: English

Economic Commission for Europe

Conference of European Statisticians

Sixty-third plenary session
Geneva, 15-17 June 2015
Item 4 of the provisional agenda
Modernisation of statistical production and services and managing for efficiency

The increasing importance of "wholesaling" data

Note by the Australian Bureau of Statistics

Summary

The paper describes the dissemination strategy of the Australian Bureau of Statistics that recognizes that presence on the web is no longer defined by the primary website of the national statistical office, but by the visibility and agility of the statistical office across multiple channels. The paper explores what 'wholesaling' data means, and how the Australian Bureau of Statistics and many other organizations are benefiting from the ability to interface with the official statistical data stores to pull data into their applications.

The paper is presented for discussion to the second session of the Conference of European Statisticians' seminar "Modernisation of statistical production and services and managing for efficiency".

GE.15-07036 (E)
1507036

Please recycle

I. Background

1. The Australian Bureau of Statistics (ABS) has a legislated function to maximise the use of data for informed decision making. ABS works within an interconnected information environment, providing statistical leadership for Australia through effective partnership with key stakeholders. Changes and innovation in the information environment are stimulating demand from government, researchers, industry and the community for access to data through rapid, flexible and secure means.

2. This paper outlines a strategy to provide better access to statistical data through data wholesaling. It explores how ABS and other organizations and individuals are benefiting from the ability to interface with the ABS data stores to pull our content into their applications using web services or Application Programming Interfaces (API).

II. Our users

3. ABS has traditionally thought about our users in terms of market segmentation, from primary school students to policy analysts, economic modellers and health researchers. More recently, ABS has focused our attention on the behaviours of our users, rather than which organization or role they perform. This has allowed us to simplify our approach to providing our content.

4. Users can be divided into four main behavioural groups so we can target our approaches to meet their requirements. These users say:

- Connect me to a number, so I can quickly tell others about it; or
- Connect me to an interesting story, so I learn something new. I would like a story or an interactive tool or game like Census spotlight (spotlight.abs.gov.au); or
- Connect me to a summary, as I want to learn more about 'Where I live' or 'What are the important health issues of today'; or
- Connect me to detailed data so I can analyse, re-use and combine it with other data, to help me or my own users make decisions.

III. The digital economy

5. Over the past decades, organizations have relied on static web sites, desktop browsers and monolithic systems to deliver content. Our traditional model is a website which delivers regular static publications and data at prearranged and predictable times. Detailed data sets are available to users in excel format, but only through manual download from the website.

6. The efficacy of this model has been disrupted by rapid technological change. The complexity of rapid changes in technologies, processing power, digital media and digital community poses a range of challenges. We know that:

- There is an exponential growth in data, but a massive reduction in the cost of holding or transmitting data. This brings data management and distribution challenges;
- The value chain has shifted from providers to aggregators or data intermediaries such as new media platforms. A challenge is how to attract attention or ensure attribution (and thus quality assurance) where content creation and distribution channels aren't limited to traditional means;

- The capacity to interact with new media has grown beyond traditional rigid models. Users can instantly access content, share, comment on, and query content at any time of the day;
- Speed is everything in the digital economy. Milliseconds mean money. Everyone wants information at the tip of their fingers, fast and efficiently;
- The applications development community is a significant player. Through the use of new web technologies, new applications can be constructed quickly and cheaply, to meet broad or specific needs;
- Information can be readily and automatically interpreted by machines, so machines can perform more of the tedious work involved in finding, combining, and acting upon information on the web;
- There is an increasing use of artificial intelligence systems within the Internet, enabling automatic decision making.

7. Our key users are players in, and impacted by, the digital revolution. Government, business and researchers need faster and better access to data. They want access to detailed fully described macro and confidentialised micro data. They want to integrate data across traditional and non-traditional statistical subject matter boundaries and from different sources to address complex problems.

8. They do not want to interact with the traditional model. Manually downloading data files from a website is inefficient, slow and unproductive, and impedes innovation. Our output is often an important input to their business processes.

9. Some users expect us to supply data using agreed data structure definitions in formats such as the Statistical Data and Metadata Exchange (SDMX). This is effective for both parties and allows them to 'pull' or access our information on the fly. Ultimately, agreement to standards for metadata and data across the statistical industry will allow users to access multiple data sources more easily and consistently, regardless of source.

IV. Transforming output and communication

10. In developing a new approach to dissemination, ABS is using the following principles to guide us to:

- Enhance coherence through storing the most detailed and fully described multi-dimensional data in an output warehouse as the 'one source of truth';
- Promote and use standard approaches to metadata, not just within the organization but across the statistical and information industries;
- Collaborate with similar organizations to share approaches and tools;
- Assemble the many varieties of potential outputs from the 'one source of truth';
- Dynamically compile content to meet the needs of different levels of statistical literacy. Whilst there will be some static content, increasingly, content will be derived dynamically from underlying detailed data and metadata;
- Assemble views of the data and metadata, through a standard library of geographic, graph and visualisation tools;
- Provide open, timely, inexpensive access to public sector information;
- Enable access to data, including large volume data, using machine to machine services;

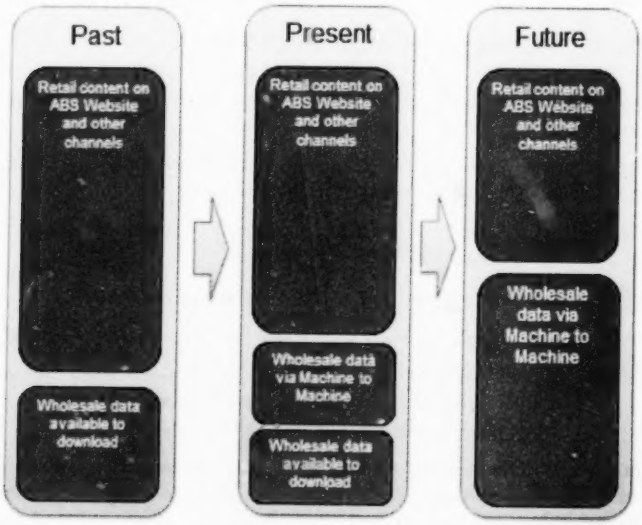
- Focus on user needs through the provision of quality, relevant and timely statistics in ways that users can easily find and understand;
- Provide appropriate content regardless of the channel or device used by users;
- Facilitate re-use of ABS data by external users and third party portals while ensuring visibility of, and attribution to, the ABS brand; and
- Ensure content designed to build statistical capability is visible alongside data, but not intrusive.

11. An important part of responding to the digital revolution is the wholesaling of our data. By wholesaling, we mean the provision of access to large quantities of data/content through machine-to-machine services to be re-used by others in the production of value added products and services for themselves, or their own users.

12. This does not mean that ABS will cease providing data, content or applications itself. We have strong community service obligations, and will always provide a comprehensive and broad range of data and summary content to the Australian community. ABS has the capabilities and the demand to provide trusted, integrated statistical solutions to customers.

13. However, what it does mean is that we will increasingly provide machine-to-machine mechanisms for ourselves and others to integrate, aggregate and re-use our more detailed data in formats and innovative solutions that ABS itself would not otherwise provide. This approach allows others to tailor their use of our data for their own specific purposes, including use within their own systems and business processes, use with commonly used commercial analytical tools, and use to deliver their own value added services.

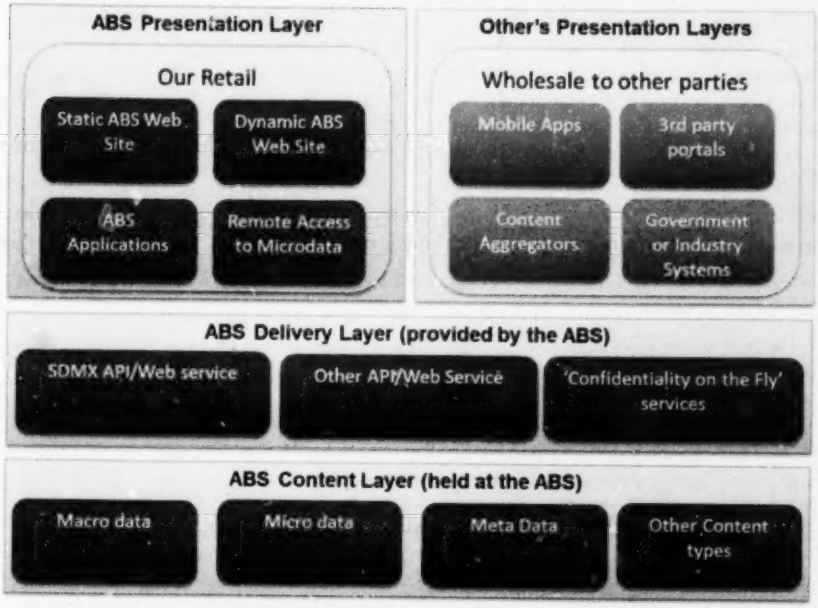
Figure 1
Transformation of ABS output


Past
Retail content on ABS Website and other channels
Wholesale data available to download
Present
Retail content on ABS Website and other channels
Wholesale data via Machine to Machine
Wholesale data available to download
Future
Retail content on ABS Website and other channels
Wholesale data via Machine to Machine

A. Flexible architecture for dynamic publishing

14. The approach requires an architecture (including business, systems and information architecture) which has separate layers for content, for delivery and for presentation.

Figure 2
The three layer model to support wholesale data



15. ABS has to be able to build, catalogue, describe, store, protect and access the data content in a consistent manner, independent of how our users will eventually use and consume it. Instead of creating traditional publications, where summary data, detailed data, analysis and stories about the data may exist in one monolithic and static publication, we will focus on creating the underlying elements of content.

16. The delivery layer connects the content to the presentation of this content. There are two important aspects of the delivery layer. The first is the tools that allow users and systems access to the content at various levels of granularity, and the second is the services, also known as APIs or web services, that will ensure our most detailed content can be transformed and made confidential before presenting it to our users ('confidentiality on the fly'). With these, assuming appropriate metadata, the content can be sliced, diced and combined with other content in third party channels.

17. The presentation layer provides a way to effectively address users' different needs, by presenting our content in many different ways, in different formats and on different devices. The presentation layer does not only include the retail provision of statistics and content on ABS channels, but also those presentations of the data and content that the ABS has wholesaled to other parties. It will encompass an increasing number of general and tailored applications, provided by us or by others.

V. Our own use of data wholesaling

18. An example of our own use of data wholesaling is the ABS Data by Region (DbR) application (www.abs.gov.au/databyregion).

19. Users want to find information about the places in which they live. They want to compare their region with others to see how it fits into the broader Australian context. Over the last decade, the ABS' National Regional Profile (NRP) has been one of the mechanisms to provide this information. The NRP provides regional data from multiple data sets into a single consistent time series. In the past, we have manually compiled and validated thousands of region specific web pages to construct the NRP. Whilst this has saved users from having to find and explore each data set themselves, this process took months, was resource intensive and inefficient. Data in the NRP, once released, wasn't updated until the next annual release, even though new regional data becomes available on a regular basis during the year.

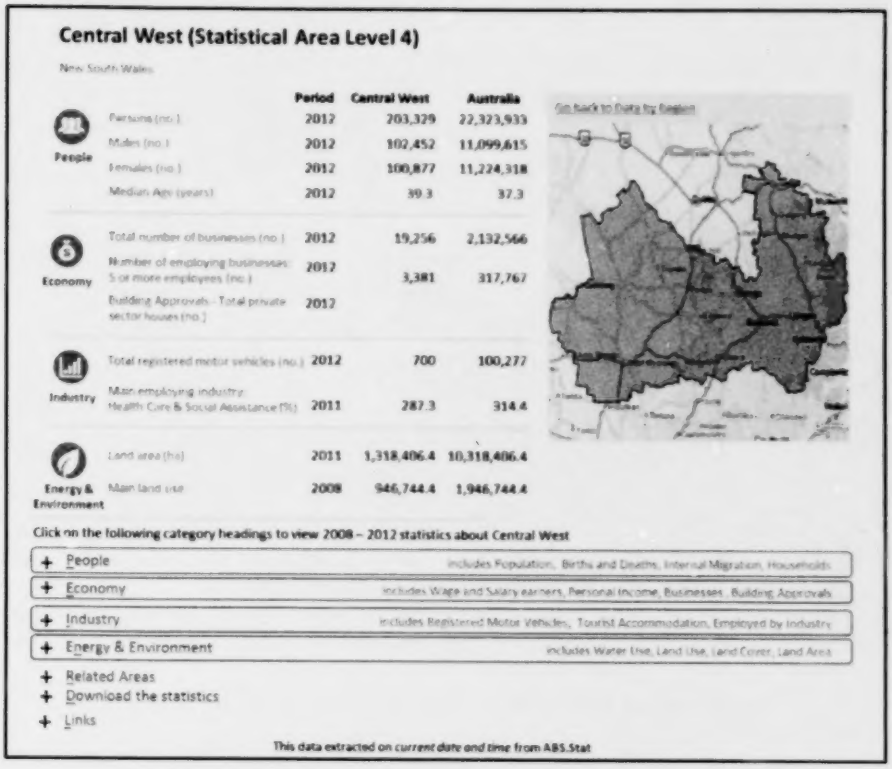
20. In July 2014, ABS developed the Data by Region application. This uses all three layers to deliver a new innovative service to our users:

- A content layer delivered through datasets in ABS.Stat (which is based on the OECD.Stat product used by a number of official statistics organizations);
- A delivery layer delivered through ABS.Stat web services and an ABS developed web service; and
- A presentation layer delivered by the Data by Region application.

21. DbR provides a single entry point for access to ABS statistics by region. It has intuitive navigation through a mapping tool and delivers new statistics as they are released. The biggest revolution of the DbR project is 'under the hood' - with the way it calls information/data from a single centralised repository. When a user selects a region to view, the data is pulled dynamically from the repository. We no longer need to manually build individual datasets for each region. This approach is effectively wholesaling our own data to our own retail applications. It will help us build a sustainable and flexible approach to the dissemination of our content.

Figure 3

Data on the Central West of New South Wales (NSW) (from Data by Region application)



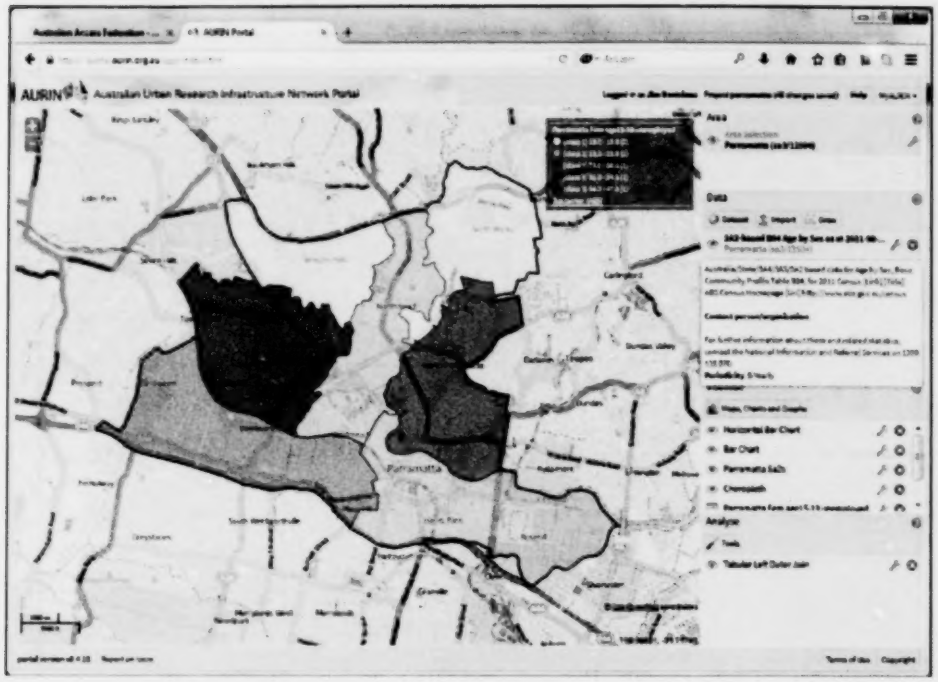
VI. Other users of our data wholesaling

22. The new ABS web services also allow other parties to access data, securely stored at ABS, in real time, to integrate this data with other datasets and present the output into their own systems. An example of this is the Australian Urban Research Infrastructure Network (AURIN), which is used across the whole of the Australian university sector and beyond (See http://aurin.org.au/).

23. AURIN's aim is to empower urban settlements research and decision making by developing e-infrastructure that drives and anticipates innovation, new methods and practice. AURIN and ABS commenced a joint project mid-2013 to make ABS data sets available to AURIN via on line and real time machine-to-machine services.

24. The project has been successful in demonstrating machine-to-machine access to 2011 Census data through a federated data hub model for delivering statistical data across Australia. AURIN has used their knowledge and tools to present this data in a sophisticated user interface that ABS would never have built. In essence, ABS has wholesaled data to AURIN, to produce a new value added product and service (the AURIN portal).

Figure 4
Screen shot from the AURIN portal that uses ABS data



25. The wholesaling of data by ABS has delivered a range of benefits to both AURIN and ABS, including:

- Data presented by AURIN is always the most recent data from the ABS;
- Re-usable services. AURIN is an urban infrastructure portal. Our web services technologies could just as easily support a dedicated environmental or economic portal;
- Productivity gains to AURIN users;
- The ABS has not needed to develop the presentation layer;
- Multiple different developers have provided analytical techniques (both existing and new) for use through the portal;
- Reinforced the need for well-described data and metadata;
- Tested aspects of our future dissemination strategy;
- Increased experience working with the growing sector of information intermediaries.

VII. Wholesaling microdata

26. A final area where we have made progress is the dynamic generation of aggregates from confidential unit record data. We have invested in the development of a suite of products that allow external research and policy users access to data for subpopulations

without jeopardising privacy and confidentiality. These products allow users to tabulate and analyse microdata and generate dynamically confidentialised outputs.

27. An example of one of these web-based products is TableBuilder (See Survey TableBuilder). This incorporates a perturbation method as part of the delivery layer for automatically protecting tables of count data and key summary statistics such as custom ranges, totals, means, medians and quintiles. This perturbation method mitigates disclosure risks from requests for similar tables, repeated requests for identical tables, and repeated requests for the same table cell within different tables. TableBuilder incorporates other confidentiality algorithms and protections that run on the fly and minimise utility loss, while ensuring that the outputs produced are not likely to enable the identification of a respondent.

28. At the moment, ABS provides the presentation layer for the confidentiality services and analytical techniques are restricted to those we are able to provide. In the future, however, we plan to provide the confidentialisation 'on the fly' as a service so that others can generate confidentialised output for use within their own systems. Along with approaches to provide non-systems solutions for access to microdata, these 'confidentiality on the fly' services will open up more opportunities for our users, especially researchers, to interact with and reuse our more detailed datasets.

VIII. Going forward

29. ABS recognises the user is no longer just a 'reader' of the Web. Today's web user is a 'protagonist' of the Web. They increasingly play and want to play an active role to disseminate, to share, to discuss, to promote and to reuse information. Wholesaling opens up a wide range of opportunities to connect into the digital economy. By providing wholesale access to data and metadata, users will be able to draw ABS content into their own preferred streams of content, whether that be through new media, their own systems or other as yet unknown innovations. Whilst there will always be a need for ABS to provide content to meet the needs of our users, we predict that the demand for 'wholesaling' of data will continue to grow.

30. We are experts at collecting, processing and analysing statistics. We are not always experts at presenting that data, nor do we have the resources to provide our data in every format or creative style that our users want. We will need others to do that.

31. There are a number of challenges:

- With increased re-use, mashing and distribution of data, sophisticated techniques need to be developed to keep track of attribution, ensure data provenance and protect our brand;
- Our processes and infrastructure need to be scaled to deliver consistent reliability and performance;
- These new approaches present both opportunities and challenges and we need to ensure they mesh well with our strategic goals, operational model and protect our organization's viability.

32. Small steps or large, ABS is committed to this transformation as one worth doing and recognises many other national statistical offices and international agencies are making similar changes. Wholesaling of data is an important part of modernising and maximising the use of ABS data.

References

Branson, M 2014, *Measuring Interest Group Expectations and Trust*, paper presented to 2014 IAOS Conference, Da Nang, Vietnam, 8-10 October 2014.

Borowik, J., Branson, M. and Watson, D. 2013, *Interactive Content and Dynamic Publishing – A vital part of an NSO'S output and communications strategy*, paper presented at IASE, Macau 2013.

Montgomery, T 2014, *Dynamic Publishing – how far can we go?*, paper presented at 2014 IMAODBC Conference, Evora, Portugal, September 2014